JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 46, 447–462 (1974)

On Random Least-Square Analysis*

J. W. DOWDEE† AND T. T. SOONG

Department of Engineering Science, State University of New York at Buffalo, Buffalo, New York 14214

Submitted by Richard Bellman

Dispersion in the output data of a system can be analyzed as either noise fluctuations about a deterministic model or as the noise with added fluctuations due to randomness in the model itself. This latter interpretation finds applications in the identification of inherently random systems which provide rational models for systems such as biological and economic systems. It is shown that the computational procedure is closely related to traditional least-square analysis. Both linear and nonlinear models are considered. Results of computer simulations are presented for some simple cases.

1. INTRODUCTION

Least-square analysis has been used extensively to estimate the parameters associated with linear models. In this analysis, the dispersion in the observed data is modeled as random noise fluctuations about a deterministic function. This dispersion, however, could arise, not from measurement noise, but from the fact that the model itself is randomly varying. This latter interpretation is especially important when modeling inherently random systems such as biological or economic systems. For example, in the compartmental analysis of biological systems, uncertainties in rate constants arise from environmental effects and the variation of patient parameters. The output of such a system is therefore subjected to random fluctuations which cannot be adequately modeled by a strictly measurement noise model. An appropriate model must also take into account the randomness of the system itself.

This paper presents a procedure for the analysis of data associated with randomly varying models. The model to be discussed represents data from p experiments and is of the form

$$y_k(t_j) = \sum_{i=1}^{m} a_{ik}(t_j)\, g_i(t_j) + \epsilon_{kj}\,, \qquad k = 1, 2, \ldots, p_j\,, \quad j = 1, 2, \ldots, N, \tag{1}$$

* Supported by the National Science Foundation under Grants GK-1834X and GK-31229.

† Presently with TRW Systems, Redondo Beach, CA.

where $g_i(t_j)$, $i = 1,..., m$, are known deterministic functions evaluated at time t_j, $a_{ik}(t_j)$, $i = 1, 2,..., m$, are unknown parameters which in general may depend on the experiment k and the time t_j, $y_k(t_j)$ is the observed response from the kth experiment at time t_j, p_j is the number of observations at t_j, and ϵ_{kj} is the random error associated with the kth experiment and the jth time point.

Based on the properties attributed to $a_{ik}(t_j)$, several classifications of the above model can be defined as follows:

(a) Ordinary Least-Square (OLS) Model. In this classification the unknown parameters $a_{ik}(t_j)$ are represented by

$$a_{ik}(t_j) = a_i, \qquad i = 1,..., m \tag{2}$$

for all k and j, where the a_i are deterministic constants. The only randomness in this model is due to the measurement errors ϵ_{kj}.

(b) Constant Coefficient Random Least-Square (CCRLS) Model. For this model, one has

$$a_{ik}(t_j) = a_{ik}, \qquad i = 1,..., m, \quad k = 1,..., p_j \tag{3}$$

for all t_j, where a_{ik} is the realization of a random variable. In this model, the unknown coefficients are deterministic for a given experiment but vary randomly from experiment to experiment.

(c) Time-Varying Coefficient Random Least-Squares (TVCRLS) Model. In this model, the coefficients $a_{ik}(t_j)$ vary randomly with i, k, and t_j. The random measurement error ϵ_{kj} may be deleted from this model since it cannot be distinguished from an unknown coefficient.

A subclassification of practical importance is one where there is only one experiment, i.e., $p_j = 1$ for all j. This subclassification will be referred to as the *unreplicated data case.*

The OLS model has received considerable attention beginning with Gauss [3] and Markoff [6]. Comprehensive lists of references can be found in Rao [9] and Draper and Smith [1]. Rao [9] considered CCRLS models and Swamy [14] extended Rao's results and applied the analysis to study the behavior of economic systems. Fisk [2] and Hussain [5] also considered CCRLS models in the context of experimental design. The TVCRLS models with unreplicated data has previously been considered by Hildreth and Houck [4] from a least-square point of view and by Nelder [7] from a likelihood-function viewpoint. This paper considers the TVCRLS model and extends the analysis of Hildreth and Houck for the unreplicated data case as well as develops the theory applicable to the case of multiple observations at each

time point. This analysis is presented in Sections 2.1 and 2.2 for linear models and in Section 2.3 for a simple nonlinear model. Section 3 presents some numerical results illustrating the theory of Section 2.

2. TVCRLS Model Analysis

The coefficients $a_{ik}(t_j)$ in Eq. (1) represent the kth realizations of the random variables a_i at time t_j and are assumed to satisfy the following properties:

(a) $E[a_{ik}(t_j)] = \mu_i$ for all k and t_j,

(b) $\operatorname{cov}[a_{ik}(t_j)\, a_{\ell q}(t_r)] = \begin{cases} \sigma_{i\ell}\delta_{jr}\delta_{kq} & i \neq \ell, \\ \sigma_i^2\delta_{jr}\delta_{kq} & i = \ell, \end{cases}$

where δ_{jr} is the Kronecker delta function.

(c) $A = [a_1, a_2, \ldots, a_m]^T$ is a Gaussian random vector with mean $\alpha_1 = [\mu_1, \mu_2, \ldots, \mu_m]^T$ and covariance matrix $C = \{c_{ij}\}$, $i, j = 1, 2, \ldots, m$, where $c_{ij} = \sigma_{ij}$, $i \neq j$, and $c_{ii} = \sigma_i^2$.

It is also assumed that $N \geqslant m(m+1)/2$.

At each t_j, the observation $y_k(t_j)$ therefore represents the kth realization from a Gaussian random variable $y(t_j)$ with mean

$$\mu_y(t_j) = G_j^T \alpha_1 \tag{4}$$

and variance

$$\sigma_y^2(t_j) = G_j^T C G_j, \tag{5}$$

where

$$G_j = [g_1(t_j), \ldots, g_m(t_j)]^T.$$

2.1. *Replicated Data Case*

For the replicated data case it will also be assumed that $p_j > 1$ for all time points t_j, $j = 1, 2, \ldots, N$. In view of the assumptions (a)–(c), an unbiased estimate of $\mu_y(t_j)$ is

$$\hat{\mu}_y(t_j) = \frac{1}{p_j} \sum_{k=1}^{p_j} y_k(t_j). \tag{6}$$

Defining $Y_D = [\hat{\mu}_y(t_1), \ldots, \hat{\mu}_y(t_N)]^T$, $B_1 = [G_1, \ldots, G_N]^T$, and $\epsilon_1 = Y_D - B_1\alpha_1$, it is easily shown that $E(\epsilon_1) = 0$ and

$$\operatorname{cov}(\epsilon_1) = \Sigma_1 = \operatorname{diag}\left[\frac{\sigma_y^2(t_1)}{p_1}, \ldots, \frac{\sigma_y^2(t_N)}{p_N}\right]. \tag{7}$$

Equation (4) can now be replaced by

$$Y_D = B_1\alpha_1 + \epsilon_1 , \tag{8}$$

and, based upon ordinary least-squares, an estimate $\hat{\alpha}_1$ of α_1 is given by

$$\hat{\alpha}_1 = (B_1{}^T\Sigma_1^{-1}B_1)^{-1}B_1{}^T\Sigma_1^{-1}Y_D . \tag{9}$$

The weight matrix Σ_1^{-1} is generally not known *a priori* and must also be estimated from the data.

To proceed, let us first cite a result in matrix theory [13]:

THEOREM 1. *If A is an $n \times 1$ matrix, B is $n \times n$, and C is $n \times n$, then*

$$A^TB^TCBA = [\text{vec}(B^T)]^T\,[C \otimes (AA^T)]\,\text{vec}(B^T), \tag{10}$$

where $\otimes$ is the Kronecker product [8], *and* $\text{vec}(\cdot)$ *is the vector operator defined by* $\text{vec}(B) = [b_1{}^T,\ldots, b_m{}^T]^T$, *where b_j is the j*th *column of B.*

COROLLARY 1. *If A is $m \times 1$ and C is $m \times m$, then*

$$A^TCA = [\text{vec}(AA^T)]^T\,\text{vec}(C), \tag{11}$$

which is a direct consequence of Theorem 1 *by letting B be the identity matrix.*

Using this corollary in Eq. (5) yields

$$\sigma_y{}^2(t_j) = [\text{vec}(G_jG_j{}^T)]^T\,\text{vec}(C). \tag{12}$$

Since C is a covariance matrix, $m(m-1)/2$ elements of $\text{vec}(C)$ are identically equal. To avoid problems with a singular coefficient matrix, the following definitions are introduced.

DEFINITION 1. Let C be a $m \times m$ symmetric matrix. The $m(m+1)/2 \times 1$ column vector $\text{utvec}(C)$ is defined as

$$\begin{aligned}\text{utvec}(C) = [&C_{11}, C_{12},\ldots, C_{1m}, C_{22},\ldots, C_{2m},\\ &\times C_{33},\ldots, C_{m-1,m-1}, C_{m-1,m}, C_{m,m}],\end{aligned}$$

i.e., by forming a vector with the upper triangular terms of C.

DEFINITION 2. If C is an $n \times n$ symmetric matrix, $\text{dutvec}(C)$ is an $n(n+1)/2 \times 1$ vector formed by replacing c_{ij}, $i \neq j$, in $\text{utvec}(C)$ by $2C_{ij}$.

With these definitions, Eq. (12) can be written as

$$V = B_2\alpha_2 , \tag{13}$$

where

$$V = [\sigma_y^2(t_1),..., \sigma_y^2(t_N)]^T, \qquad \alpha_2 = \text{utvec}(C),$$

and

$$B_2 = \{[\text{dutvec}(G_1 G_1^T)],..., [\text{dutvec}(G_N G_N^T)]\}^T.$$

The variances $\sigma_y^2(t_j)$ can be estimated as

$$s_y^2(t_j) = \frac{1}{p_j - 1} \sum_{k=1}^{p_j} [y_k(t_j) - \hat{\mu}_Y(t_j)]^2, \tag{14}$$

where, by assumptions (a)–(c), the estimator is unbiased. Defining

$$S = [S_y^2(t_1),..., S_y^2(t_N)]^T \qquad \text{and} \qquad \epsilon_2^2 = S - B_2\alpha_2,$$

it is easily shown that

$$\begin{aligned} E(\epsilon_2) &= 0, \\ \text{cov}(\epsilon_2) &= \Sigma_2 = \text{diag}\left[\frac{2\sigma_y^4(t_1)}{p_1 - 1},..., \frac{2\sigma_y^4(t_N)}{p_N - 1}\right]. \end{aligned} \tag{15}$$

Equation (13) can be rewritten as

$$S = B_2\alpha_2 + \epsilon_2, \tag{16}$$

and ordinary least-squares leads to an estimate $\hat{\alpha}_2$ of α_2 in the form

$$\hat{\alpha}_2 = (B_2^T \Sigma_2^{-1} B_2)^{-1} B_2^T \Sigma_2^{-1} S. \tag{17}$$

As was the case with Σ_1^{-1}, the weight matrix Σ_2^{-1} involves knowledge of the covariance matric C which is a function of the unknown vector α_2; this is indicated explicitly by denoting Σ_1^{-1} and Σ_2^{-1} as $\Sigma_1^{-1}(\alpha_2)$ and $\Sigma_2^{-1}(\alpha_2)$, respectively. As noted by Soong [12], this difficulty can be circumvented by using $\hat{\Sigma}_1^{-1} = \Sigma_1^{-1}(\hat{\alpha}_2)$ and $\hat{\Sigma}_2^{-1} = \Sigma_2^{-1}(\hat{\alpha}_2)$ as the best possible approximations of Σ_1^{-1} and Σ_2^{-1}. Equations (9) and (17) can now be written in the form

$$\hat{\alpha}_1 = (B_1^T \hat{\Sigma}_1^{-1} B_1)^{-1} B_1^T \hat{\Sigma}_1^{-1} Y_D, \tag{18}$$

$$\hat{\alpha}_2 = (B_2^T \hat{\Sigma}_2^{-1} B_2)^{-1} B_2^T \hat{\Sigma}_2^{-1} S, \tag{19}$$

which represent two nonlinear simultaneous vector equations in two vector unknowns. Several techniques are available for solving these equations. A simple procedure which has provided good results is an iteration method based on successive substitution. To initialize the procedure, an initial estimate $\hat{\alpha}_{2[0]}$ of α_2 must be obtained. As indicated in Section 3, the unweighted least-square estimate

$$\hat{\alpha}_{2[o]} = (B_2^T B_2)^{-1} B_2^T S$$

has successfully provided starting values for a wide variety of cases. Once the procedure is initialized, the qth estimates of α_1 and α_2 are obtained as

$$\hat{\alpha}_{2[q]} = (B_2{}^T \hat{\Sigma}_{2[q-1]}^{-1} B_2)^{-1} B_2{}^T \hat{\Sigma}_{2[q-1]}^{-1} S, \tag{20}$$

$$\hat{\alpha}_{1[q]} = (B_1{}^T \hat{\Sigma}_{1[q]}^{-1} B_1)^{-1} B_1{}^T \hat{\Sigma}_{1[q]}^{-1} Y_D\,, \tag{21}$$

where

$$\hat{\Sigma}_{1[q]}^{-1} = \Sigma_1^{-1}(\hat{\alpha}_{2[q]}) \qquad \text{and} \qquad \hat{\Sigma}_{2[q-1]}^{-1} = \Sigma_2^{-1}(\hat{\alpha}_{2[q-1]}).$$

The estimates obtained by the above procedure are unbiased and, if the weight matrices Σ_1^{-1} and Σ_2^{-1} were known, the estimates would have minimum variance among all linear estimators. Rao [11] has shown that if the weight matrices are not known, there exists the possibility of the unweighted least-square estimates being more efficient than the estimates using estimated weights. This problem can be studied analytically by defining

$$\Lambda_i = (B_i{}^T B_i)^{-1} B_i{}^T \Sigma_i Z_i\,, \qquad i = 1, 2, \tag{22}$$

where $r_i = \text{rank}(B_i)$ and the Z_i are $N \times (N - r_i)$ matrices such that $B_i{}^T Z_i = 0$, $i = 1, 2$. Using these definitions, Rao shows that the unweighted least-squares procedure will provide the better estimates only if the Λ_i are near zero. Since in practical problems, the covariance matrices Σ_1 and Σ_2 will differ substantially from the identity matrix, it is expected that using estimated weights will provide better results than the use of equal weighting. This expectation has been substantiated by the numerical results presented in Section 3.

A case of special interest and simplicity occurs when $p_j = p$ for all j and it is known *a priori* that $C = \sigma^2 I$. Since only the relative magnitudes of the weights are important, the weight matrices for this case can be calculated as

$$\Sigma_2^{-1} = (\Sigma_1^{-1})^2 = \text{diag}\left[\frac{1}{(G_1 G_1{}^T)^2}, \ldots, \frac{1}{(G_N G_N{}^T)^2}\right]. \tag{23}$$

The weight martices are therefore known and Eqs. (9) and (17) can be used directly to obtain $\hat{\alpha}_1$ and $\hat{\alpha}_2$ without the need of iteration. Hypothesis test procedures can be used to determine when $C = \sigma^2 I$ is a reasonable assumption.

2.2. *Unreplicated Data Case*

When $p_j = 1$ for all j, Eq. (18) remains valid but Eq. (19) must be modified since S is no longer defined. In this case, it is noted that

$$r_j{}^2 = [y(t_j) - \mu_j]^2, \qquad j = 1, 2, \ldots, N \tag{24}$$

is an unbiased estimator of $\sigma_y^2(t_j)$ with a variance equal to $2\sigma_y^4(t_j)$. Since μ_j is not known, α_1 is first estimated and $\hat{\mu}_j$ calculated as

$$\hat{\mu}_j = G_j^T \hat{\alpha}_1 . \tag{25}$$

Using this estimate, $\hat{r}_j^2$ can be approximated as

$$\hat{r}_j^2 = [y(t_j) - \hat{\mu}_j]^2, \qquad j = 1, 2, \ldots, N, \tag{26}$$

and the vector $\hat{R}^{(2)} = [\hat{r}_1^2, \ldots, \hat{r}_N^2]^T$ can be used instead of S in Eq. (19). The weight matrix Σ_2^{-1} must also be replaced by

$$\Sigma_2^{-1} = \operatorname{diag}\left[\frac{1}{2\sigma_y^4(t_1)}, \ldots, \frac{1}{2\sigma_y^4(t_N)}\right] . \tag{27}$$

As in the replicated data case, Σ_1^{-1} and Σ_2^{-1} are not known *a priori* so an iteration procedure is required. To initialize the procedure, the initial estimates $\hat{\alpha}_{1[0]}$ and $\hat{\alpha}_{2[0]}$ can be obtained by an unweighted least-squares. They are given by

$$\hat{\alpha}_{1[o]} = (B_1^T B_1)^{-1} B_1^T Y_D , \tag{28}$$

$$\hat{R}_{[o]}^{(2)} = \{[y(t_1) - G_1^T \hat{\alpha}_{1[o]}]^2, \ldots, [y(t_N) - G_N^T \hat{\alpha}_{1[o]}]^2\}^T, \tag{29}$$

$$\hat{\alpha}_{2[o]} = (B_2^T B_2)^{-1} B_2^T \hat{R}_{[o]}^{(2)} . \tag{30}$$

After initialization, successive substitution can be used to obtain the estimates of α_1 and α_2. As in the replicated data case, if it is known a priori that $C = \sigma^2 I$, then an iteration procedure is not required.

The estimate $\hat{\alpha}_2$ obtained by the above procedure is not unbiased since $E[\hat{R}^{(2)}] \neq B_2\alpha_2$. Noting that

$$E[\hat{r}_j^2] = \sigma_y^2(t_j) + 2\operatorname{cov}\{[y(t_j) - \mu_j], [\hat{\mu}_j - \mu_j]\} + \operatorname{var}[\hat{\mu}_j], \tag{31}$$

it is seen that if $\hat{\mu}_j$ is consistent, then $\hat{\alpha}_2$ will be asymptotically unbiased.

If unbiased estimates of α_2 are desired, a modified version of a procedure suggested by Hildreth and Houck [4] can be used. The estimate of α_1 on the qth iteration, $\hat{\alpha}_{1[q]}$, is obtained, as in the above procedure as

$$\hat{\alpha}_1 = (B_1^T \hat{\Sigma}_1^{-1} B_1)^{-1} B_1^T \hat{\Sigma}_1^{-1} Y_D , \tag{32}$$

where the subscripts denoting the iteration number have been dropped for notational simplicity. The qth vector of residuals

$$\hat{R} = B_1 \alpha_1 \tag{33}$$

is then related to the data vector Y_D as

$$\hat{R} = MY_D, \tag{34}$$

where M is an $N \times N$ symmetric idempotent matrix in the form

$$M = I - B_1(B_1{}^T\hat{\Sigma}_1^{-1}B_1)^{-1}B_1{}^T\hat{\Sigma}_1^{-1}. \tag{35}$$

The residual vector has zero mean and a covariance matrix

$$\Sigma_{\hat{R}} = M\Sigma_1 M. \tag{36}$$

The diagonal terms of $\Sigma_{\hat{R}}$ leads to

$$E[\hat{r}_j{}^2] = \sum_{k=1}^{N} m_{jk}^2 \sigma_y{}^2(t_k), \tag{37}$$

where m_{jk} is the jkth element of M. Defining $M^{(2)}$ as an $N \times N$ matrix whose ijth element is m_{ij}^2, one has

$$E[\hat{R}^{(2)}] = M^{(2)}B_2\alpha_2. \tag{38}$$

With ϵ_3 defined as

$$\epsilon_3 = \hat{R}^{(2)} - M^{(2)}B_2\alpha_2, \tag{39}$$

it is easily seen that ϵ_3 has zero mean. Therefore,

$$\hat{R}^{(2)} = M^{(2)}B_2\alpha_2 + \epsilon_3 \tag{40}$$

is a linear model from which α_2 can be estimated by ordinary least-square procedures. We thus have

$$\hat{\alpha}_2 = (B_2{}^TM^{(2)}\Sigma_3^{-1}M^{(2)}B_2)^{-1}B_2{}^TM^{(2)}\Sigma_3^{-1}\hat{R}^{(2)}, \tag{41}$$

where Σ_3 is the covariance matrix of $R^{(2)}$.

From Eq. (38) it is readily shown that this estimate of α_2 is unbiased. The ijth element of the covariance matrix Σ_3 can be found as

$$\operatorname{cov}[\hat{r}_i{}^2, \hat{r}_j{}^2] = 2\sum_{k=1}^{N}\sum_{\ell=1}^{N} m_{ik}m_{i\ell}m_{jk}m_{j\ell}\sigma_y{}^2(t_k)\,\sigma_y{}^2(t_\ell). \tag{42}$$

Since these covariances depend on the unknown quantities $\sigma_y{}^2(t_k)$, $k = 1,\ldots, N$, they must also be estimated from the data and incorporated into an iteration procedure as previously described. An unweighted least-square estimate could be used to initialize the procedure.

Several practical observations should be made in comparing the unbiased with the biased estimation procedure.

(a) The coefficient matrix $M^{(2)}B_2$ in the unbiased procedure depends on Σ_1 and therefore on the estimate of α_2. This difficulty can be circumvented by using an iteration procedure but is computationally more time-consuming than the biased estimation procedure.

(b) The matrix Σ_3 is an $N \times N$ nondiagonal matrix which becomes impractical to invert for large N.

(c) As N becomes large, M approaches the identity matrix and the unbiased and biased procedures become equivalent.

2.3. *A Nonlinear Problem*

The ideas presented in Sections 2.1 and 2.2 can also be used in connection with some nonlinear problems which are amendable to linearization. Let $y_k(t_j)$, $k = 1,..., p$ and $j = 1,..., N$, represent the output of p random systems. It is assumed that the output can be modeled as

$$y_k(t_j) = a \exp[\beta_k(t_j)\, t_j], \qquad k = 1,..., p, \quad t = 1,..., N, \tag{43}$$

where a is an unknown constant and $\beta_k(t_j)$ is the kth realization of a Gaussian random variable β at time t_j. It is further assumed that

$$\left.\begin{aligned} &\text{(a)} \quad E[\beta_k(t_j)] = \mu \text{ for all } k \text{ and } t_j, \\ &\text{(b)} \quad \operatorname{cov}[\beta_k(t_j), \beta_\ell(t_r)] = \begin{cases} \sigma^2 & k = \ell, j = r, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}\right\} \tag{44}$$

In a biomedical systems context, this would describe a one-compartment system with a deterministic initial condition and a random rate constant. The problem is to estimate a, μ, and σ^2 from the output data.

Taking expected values of Eq. (43) yields

$$E[y_k(t_j)] = a \exp\left(\mu t_j + \frac{\sigma^2 t_j^2}{2}\right) \tag{45}$$

for all k. Therefore

$$\ln E[y_k(t_j)] = \ln(a) + \mu t_j + \frac{\sigma^2 t_j^2}{2}. \tag{46}$$

To put Eq. (46) in the form suitable to analysis by the procedure developed in Section 2.1, an unbiased estimate of $\ln E[y_k(t_j)]$ is required.

THEOREM 2. *Let*

$$Z_k(t_j) = \ln[y_k(t_j)], \tag{47}$$

$$\bar{Z}_j = \sum_{k=1}^{p} Z_k(t_j)/p, \tag{48}$$

$$S_Z^2(t_j) = \frac{1}{p-1} \sum_{k=1}^{p} [Z_k(t_j) - \bar{Z}_j]^2. \tag{49}$$

Then $\bar{Z}_j + \frac{1}{2}S_Z^2(t_j)$ is an unbiased estimate of $\ln E[y(t_j)]$ with variance

$$\frac{\sigma^2 t_j^2}{p} + \frac{\sigma^4 t_j^4}{2(p-1)}. \tag{50}$$

Proof. The random variable $Z(t_j) = \ln[y(t_j)]$ is Gaussian with mean

$$E[Z(t_j)] = \ln(a) + \mu t_j \tag{51}$$

and variance

$$\sigma_Z^2(t_j) = \text{var}[Z(t_j)] = \sigma^2 t_j^2. \tag{52}$$

The sample mean $\bar{Z}_j$ is therefore an unbiased estimate of $E[Z(t_j)]$ with variance $\sigma_Z^2(t_j)/p$. Similarly, the sample variance $S_Z^2(t_j)$ is an unbiased estimate of $\sigma_Z^2(t_j)$ and has a variance $2\sigma_Z^4(t_j)/(p-1)$. Since $Z(t_j)$ is Gaussian, $\bar{Z}_j$ and $S_Z^2(t_j)$ are statistically independent. Using these facts, the theorem follows directly.

Therefore, Eq. (46) can be written as

$$\bar{Z}_j + \frac{1}{2} S_Z^2(t_j) = \ln(a) + \mu t_j + \frac{t_j^2 \sigma^2}{2} + \epsilon_j \qquad j = 1,..., N, \tag{53}$$

where ϵ_j is a random variable with zero mean and variance given by (50). This model is of the form of Eq. (1) and the analysis procedure in Section (2.1) can be applied.

3. NUMERICAL RESULTS

The analysis of Section 2 will be illustrated by considering several simple problems. The linear models to be investigated are of the form

$$y_k(t_j) = a_k(t_j) + b_k(t_j)\, t_j\,, \qquad j = 1,..., N, \quad k = 1,..., p, \tag{54}$$

where the coefficients $a_k(t_j)$ and $b_k(t_j)$ satisfy assumptions (a)–(c) of Section 2. The means of the Gaussian random variables a and b will be denoted by μ_a

and μ_b, respectively, and the variances by σ_a^2 and σ_b^2. The covariance between a and b is denoted by σ_{ab} and the correlation coefficient by ρ. T will denote the length of the data span, and all data will be generated in the interval $(0, T)$.

3.1. *Sampling Policy*

The variance of $y(t_j)$ is

$$\sigma_y^2(t_j) = \sigma_a^2 + 2t_j\sigma_{ab} + \sigma_b^2 t_j^2. \tag{55}$$

Since weights need only be porportional to the reciprocal of the variances the weights used in the estimation of $\alpha_1 = [\mu_a, \mu_b]^T$ can be calculated as

$$w_j = 1/(1 + 2t_j\rho k + kt^2), \qquad j = 1,..., N, \tag{56}$$

where $k = \sigma_b^2/\sigma_a^2$. The covariance matrix associated with the estimate $\hat{\alpha}_1 = [\hat{\mu}_a, \hat{\mu}_b]^T$ is easily derived as being proportional to

$$(B_1^T \Sigma_1^{-1} B_1)^{-1} \propto \begin{bmatrix} \sum_{j=1}^{N} w_j & \sum_{j=1}^{N} t_j w_j \\ \sum_{j=1}^{N} t_j w_j & \sum_{j=1}^{N} t_j^2 w_j \end{bmatrix}^{-1}. \tag{57}$$

It is therefore observed that the variances and the covariance of the estimates will be a function of how the data points are chosen in the data interval. An examination of Eq. (57) shows that $\text{cov}(\hat{\mu}_a, \hat{\mu}_b)$ will be small if

$$T \gg 1/k^{1/2} = \sigma_a/\sigma_b$$

and half the samples are obtained in the interval $t_j \ll \sigma_a/\sigma_b$ and half in the interval $t_j \gg \sigma_a/\sigma_b$. This sampling procedure will be referred to as the boundary point sampling policy. When this policy is used, the $\text{var}[\hat{\mu}_a]$ is reduced mainly by the samples in the interval $t_j \ll \sigma_a/\sigma_b$ and $\text{var}[\hat{\mu}_b]$ is reduced by the samples in which $t_j \gg \sigma_a/\sigma_b$. In general, this sampling policy will result in lower variance estimators than the use of the more traditional equally spaced sampling procedure (this conclusion also is valid for ordinary unweighted least-squares). To be implemented in practice, this policy requires some a priori knowledge on the relative magnitudes of σ_a and σ_b.

A compromise between boundary point sampling and equally spaced sampling provides very good results when no a priori information is available. This procedure, called *mixed mode sampling*, is to sample rapidly at the beginning and at the end of the data span and to use equally spaced sampling in between these extremes.

The covariance matrix of $\hat{\alpha}_2 = [\hat{\sigma}_a^2, \hat{\sigma}_{ab}, \hat{\sigma}_b^2]^T$ is not as easy to analyze. However, numerical calculations of $(B_2^T \Sigma_2^{-1} B_2)^{-1}$ indicate that mixed mode sampling is superior to equally spaced sampling. Table I gives some typical results for various values of ρ and k. It is noted that some data points are needed in the middle of the data span in order to reduce the variance of $\hat{\sigma}_{ab}$. The variance on all the estimates increase monotonically with k.

TABLE I

Comparison of Sampling Policies ($T = 10, N = 40$)[a]

Type:	ρ	k	Variances $\hat{\mu}_a$	$\hat{\mu}_b$	$\hat{\sigma}_a^2$	$\hat{\sigma}_{ab}$	$\hat{\sigma}_b^2$
BP	−0.8	0.25	0.04	0.009	0.16	0.42	0.021
MM			0.06	0.009	0.10	0.02	0.004
ES			0.09	0.010	0.19	0.03	0.005
BP	0	1	0.05	0.05	0.19	2.12	0.23
MM			0.15	0.05	0.50	0.48	0.10
ES			0.34	0.05	1.31	0.73	0.13
BP	−0.8	4	0.06	0.24	0.22	11.18	2.53
MM			0.37	0.18	1.85	4.42	1.38
ES			1.0	0.21	8.35	9.70	2.09

[a] BP = Boundary point sampling; MM = mixed mode sampling; ES = equally spaced sampling.

In the numerical results which follow, mixed mode sampling is used exclusively.

3.2. *Replicated Data Case*

For the replicated data case, observations were generated at 40 time points with four replicates at each value of time. A summary of the results of the estimation procedure is given in Table II. The distance between the estiamted values of the parameters and the true values can be defined as

$$d_{io} = \| \hat{\alpha}_{io} - \alpha_i \|^{1/2} \qquad i = 1, 2,$$
$$d_{iR} = \| \hat{\alpha}_{iR} - \alpha_i \|^{1/2} \qquad i = 1, 2$$

where the subscript "o" refers to ordinary unweighted least-squares (OLS) estimates, "R" refers to random least-square (RLS) estimates, and $\| \cdot \|$ is the norm defined as $\| x \| = x^T x$. These distances are summarized in Table III.

From these tables, it is observed that the RLS procedure results in estimates of the mean which are about six times closer to the true values than would be

obtained with the OLS procedure. The estimates of the variances and covariance are about 24 times nearer the true values when RLS is used. This illustrates that if a system is actually random, serious degradation in results can be expected if the system is incorrectly modeled as deterministic and OLS applied.

TABLE II

Replicated Data Case (Mixed Mode Sampling, $N = 40$, $p = 4$, $T = 10$)

Parameter	True value	Ordinary least-squares	Initial estimates	Random least-squares	Iterations to convergence
μ_a	2.	0.91	0.91	1.91	4
μ_b	6.	6.35	6.35	6.12	3
σ_a^2	6.	−69.5	0.0	7.62	6
σ_{ab}	−2.	42.9	42.9	−4.86	8
σ_b^2	9	−0.72	0.0	10.46	5

TABLE III

Distance of Estimates from True Values, Replicated Data Case

Estimate	OLS	RLS
Mean vector	1.15	0.18
Variance vector	88.4	3.6

The initial estimates are obtained directly from the OLS estimates with the stipulation that if the estimate of a variance is negative, this estimate is set equal to zero. Any weights which are estimated as negative are also set to zero.

It is noted that, although the initial estimates are very poor, convergence occurs within eight iterations for all the parameters. The iteration procedure has converged for all cases studied; however, as with most numerical techniques, very bad initial estimates could cause the procedure to fail. For example, if the initial estimates of σ_a^2, σ_{ab}, and σ_b^2 were all negative, the estimated value of $\sigma_y^2(t_j)$ would be negative for all t_j. This would lead to all weights being set to zero, and the procedure could not be initialized. However, a more intelligent choice of starting values should lead to convergence.

3.3. *Unreplicated Data Case*

In the unreplicated data case, 200 observations were recorded and the biased estimation procedure used. The results of this simulation are tabulated

in Tables IV and V. As in the previous replicated data case, the RLS estimates are superior to those obtained by the OLS procedure. It is also noted that, as before, the iteration procedure converges rapidly although the initial estimates are poor.

TABLE IV

Unreplicated Data Case (Mixed Mode Sampling, $N = 200$, $T = 20$)

Parameter	True value	Ordinary least-squares	Initial estimates	Random least-squares	Iterations to convergence
μ_a	2.	1.85	1.85	2.01	4
μ_b	6.	6.11	6.11	6.06	3
σ_a^2	6.	−14.6	0.0	5.65	7
σ_{ab}	−2.	14.7	14.7	−4.48	9
σ_b^2	9.	6.08	6.08	9.81	7

TABLE V

Distance of Estimates from True Values, Unreplicated Data Case

Estimate	OLS	RLS
Mean vector	0.186	0.061
Variance vector	26.7	2.63

3.4. *Nonlinear Example*

The nonlinear model investigated has the form of Eq. (43)

$$y_k(t_j) = a \exp[\beta_k(t_j)\, t_j]$$

and satisfies the assumptions specified in Section 2.3. Data were generated at 40 time points using mixed mode sampling, and four replicates were generated at each time. The results of the simulation are summarized in Table VI.

TABLE VI

Nonlinear Problem (Mixed Mode Sampling, $N = 40$, $p = 4$, $T = 20$)

Parameter	True value	Ordinary least-squares	Initial estimate	Random least-squares	Iterations to converge
$\ln a$	1.38	13.39	13.39	1.50	1
a	4.0	654696.		4.49	
μ	−6.0	−19.44	−19.44	−6.45	2
σ^2	4.0	5.76	5.76	2.83	2

As noted from the table, the RLS estimates are superior to the OLS estimates, and the iteration procedure converges rapidly. Again, it is emphasized that if a random system is incorrectly modeled as deterministic, serious degradation in the accuracy of the estimates can result.

4. Concluding Remarks

A procedure has been presented for the analysis of data associated with inherently random systems. Both replicated and unreplicated data cases were considered. The results of this analysis indicate that the choice of modeling a system as deterministic or random will greatly influence the quality of the estimates obtained as well as the analysis procedure.

It is shown that the sampling policy may have significant effects on the accuracy of the desired estimates. If practical considerations do not dictate a particular sampling procedure, the experimental design should incorporate an optimum sampling policy.

Finally, it is worth pointing out again that for large N and unreplicated data, the biased estimation procedure is computationally more efficient than the unbiased procedure and gives essentially equivalent results.

References

1. N. R. Draper and H. Smith, "Applied Regression Analysis," John Wiley and Sons, New York, 1967.
2. P. R. Fisk, Models of the second kind in regression analysis, *J. Roy. Statist. Soc. Ser. B* **28** (1967), 266–281.
3. K. F. Gauss, "Werke" Vol. 4, Gottingen, pp. 1–93, 1809.
4. C. Hildreth and J. P. Houck, Some estimators for a linear model with random coefficients, *J. Amer. Statist. Assoc.* **60** (1968), 806–825.
5. A. Hussain, A mixed model for regressions, *Biometrika* **56** (1969), 327–336.
6. A. A. Markoff, "Wahrscheinlichkeitsrechnung," Tebner, Leipzig, 1900.
7. J. A. Nelder, Regression, model-building, and invariance, *J. Roy. Statist. Soc. Ser. A* **131** (1968), 303–315.
8. H. Neudecker, A note on Kronecker matrix products and matrix equation systems, *SIAM J. Appl. Math.* **17** (1969), 603–606.
9. C. R. Rao, "Linear Statistical Inference and Its Applications," John Wiley and Sons, New York, 1965.
10. C. R. Rao, The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves, *Biometrika* **52** (1965), 447–458.
11. C. R. Rao, Least square theory using an estimated dispersion matrix and its applications to measurement of signals, *Proceedings of Fifth Berkeley Symposium on Math. Stat. and Prob.* Univ. of Calif. Press, Los Angeles, pp. 355–371, 1967.
12. T. T. Soong, An extension of the moment method in statistical estimation, *SIAM J. Appl. Math.* **17** (1969), 560–568.

13. T. T. SOONG, Pharmacokinectics with uncertainties in rate constants—II: Sensitivity analysis and optimum dosage control, *Math. Biosci.* **13** (1972), 391–396.
14. P. A. V. B. SWAMY, "Statistical Inference in Random Coefficient Regression Models," Springer-Verlag, New York, 1971.